

CATTCGTGAATTTCAAAGGAGTGGTAAAGGGTTAAGAGGAATTTTATAACTTTGTGTACTTAATCTTTAAAGAATGGGCTGGGACTCCCAAGAG
GGTATATAGCATGTAACTTGCTTGCTTTGGCACTCCAGAGGGTGACTCTGCAGAGCAAGGGTGAGCGAGGTGGAGATATACTATACAGTCTCTCGT
CATCATCTTCAGGCTTTGGCACTCAAGTACTCCCTGGTAGGAAGTCTCTGTGAAATAACAGCATCAACATCCGATAGGAGCTGTCTCAATACAGATG
CATTTGGTGAGACAGCAAAATAGGACGAGCACTGGGGAGCTACAGCAGAATGAGGAAGAGAAGGGTCTCCAGAAACAGGGCTGGGAGATGGCTCAGT
GOTAAAGGAACCTTACTGATCTGTGCTTACATATTTCTAGCTTTGACTTCATAGCTAGTATTAAGTCTGTGCCACTTGCTTCACTGTGGCATGCT
CTCCCCATTAAGCATTAAGTACGGCTACGGCTAGGAGCATGTACATATACATTAAGCCAAATAACACACAGAAATAAGATTTTAAAGAGTTTCAGGCGAAGCA
TGTTGTCTACATACCTTTTAATTAACACACTTACGAGCTAGGAGCGAGGCAAGTGTGTGTGTGTATAGGATCTGGGCACTCCCAAGATACATAAA
TGAGACTCTGCACAAACCAAAAGAAATTTTCAAGGCGCTGGTGGCCGACGCTTTAATCTTCAGCACTCGGGAGGAGAGGAGAGTAAATTTCTGATGT
TGAGGCGACGGCTGGTCTACAGAGTGAATTTTCAAGGAGCCAGGCGCTACACAGAGAAACCTGTCTCAAATAACCAAAACCAAAACCAACAA
ACAAACCAAGAAATTTTCTGAAGCAGCAGAAATTAATACAGACCTCTCATAGAGCTTCAGGCACTTGCCCTTTGTAAGCGAAATTAAGACAGACAT
AATCATAAATTTGTTTTTGGAGATTTGACTGGCCCTGGAACTACAGCTGCTGAAACATACAGAGATCTCTCGCTCGACGCTTTTGAGCCCTAGCT
TAAGGGTATCCCTGGCTTAATTTAAGAAATTAAGAACCGGGTTAAAAATGTTGTGTAGAGAATTTGTGAAGTAAGAGAAAGTTAGCATATTTGATPAA
GCMTCAAAAGTGAATATAGCTTGAAGAAATAAGATGACAAATTAAGCAAGTTTGAATGCAGTGAGACAGGAGCTAAGTGACCTGGGAAGTAGGTGAGA
GCAAGCTAACTGAACGATGTTGAGAGAGGAGTGACAGAGCTCAGCCAGAGGCTCAGAACACGCGCTGCCAGCCAGACATGGTGCCCTGGATACCG
AGAAAGCGGGAAGCGAGCTGCTGAAAGGTCACATCAAGACCACTTCAGAACTGAGAAAGCAGGAGCTGTAATGAGGAAGGCTCAGACAGACAG
GGACAGACAGCTCAACAGCAAGCTTCTAGCTCTTGATGTAGTGAGTCAACAGATGGGAAGATGGATCTGAAACAGCTTCCAAATGGGGGTCAACCCCTAGG
AAGCCGCGCGGAAAGCAGGTTTCTCATGGAGTAGGGGAGCGAGGCTGGAGTAAACACACCAATGTGTGAAATCTTAAACAGCATATCCGAGAGTGG
AGCCACAGCAAAATGTTGAACAGAGAAGGCAAAATTAATTTGTAGATACCTCTCAACCTCCATCTTTTAAAGAAATTTGAGAGAAATTTGAGAG
CTAAATAAATTTTAAAGCAGATTTTAAAGAGGAAJAACTCCAGAAAGAGATTTTAAAGAAAGATTTTAAAGAAAGATTTTAAAGAAAGATTTGAGAGAGG
CACAGCAGTGTGGAGATGATCTTGTCTATGCTGTAGAACACAGACGACTGGCCAGTACTATATAAGCAACTGGGATGATGCTATCAAATGTTA
TAGGCTCGAAATTTGCTATCCGACAGCACTCTGCATACATATAAGCAGTAGTCTGAACTATATGCCAGAGGCTTAGCTTCTCATGTCTCAGTGA
AACAACTCATATGTGCAATCTTCTAGTTCAGCTGGCTACAGGCTCAGAGTCTTCAAGTATATGATTTTATGGGAAGCATGTGACAGATCCAGGCA
GCAGTGGCAATGGAGGAGCTGAGTGTCTATCTCTGTCTTCAAGCCACTAGGAGAGCTGACTTCAGGAGGCTAAGATGAGGGTCTTCAGGCGATGC
CCATAGACACACACTCTCTCAACAGCAGCACTCATATAGTGCACCTCCCTGGGCGAAGCATTTCAAATCTCATATCTCTGTATAACACTAGTGT
TATCAGGTGACATAAATAGTACGTACATCACTAGCTGCTTGTATTCTAAATCTCAGAGGGAATGGCTAAGAGCTTAAAGAGCTCCACAGGAGAG
CTTGCTCTATAGATGTCAAGATTTTAAATAATAGTCAACAGCAATTAAGATGATCAATCAAGGGTAGTCAAGTGAACACAGATAGAGATCCAGAA
TTAAGTCACTATGATATGATATCTGATTTTATCAGAGGTGAAATTTTAACTATGGCTTTGTGTGTGAGAGCTCAAGTGTATGCTCTGGTCTGCTG
GAACCTCATATGTAGACAGGGCTGGCTGCTTTAACTCAGAGATCCACAGCTCTGGGATTAAGAGTATGGCCACAGCTCTGGCCCAATATGGCAT
TAAATATCTTTGTAGATATCTACTTTTGAAGAAAGTAAATCCTGTGGATTTATTTTGTGATCCAAATACCTATTTTGTGACTATTTTCACTCT
GATCTGAAAGGTGGCGAGGCTGGGGTGTGCTCTGAGCTGCAATCTTGCTAGCATCTTGCTAGCATCTGGCTGGCTCTCTAGCATCTTCTACATCACA
GCAGACAGGCGCAGTAAACGACCAAGCATGATGAATCTTCCTCTCCAGCCCTACAGACTGGGACAGAGAACACAGCTGAGGCGGCTGGGG
TTATGTATGTAGTGGAGTGTAGCTGATGTTCTATAGAGAGCCCTGTCTCAAGAACAAACAGCAGAGATTTTAAAGCAATTAATGTAGAGGAGT
GGAGTGGCTGTGGTGTGAAGACACTGGCTCTCTTCAGAGATCTTGAGTCTAAATCCGACAGCAACCATGGTGGCTGCAACCACTCTATAAGGG
GATGATCCCTCTCTCTGGTGTCTGGAAGCAGCTCTCATACATAAATAAATGAATATCTTTAAATAAATAATGAATATCTTGAATTAAGAGTGA
TTTGTAGATATGATATAACAGAAATACCTTGATATCTGCAATGGAAAAATCTCTTAAACGGGATGCAGAAATCTAGGCCACAGAGAAATTA
TTAATAAATAGATAAATTTGAAGAGCAGAGCTGGGAGAGATGGCTCAGTGTAGAGACATCACTGAGTGAACACAGATAGAGATCCAGAA
ACTCTGGTTCCAGGGCTTAACTATGAGAGCTCTCTGCTGTCTGAGAGCACTGTAGTCACTCAGTCACTGAGTCAAGTGAATTAACACTATGAT
TACAAAGTTCCAGAGCGCAGGCTACACAGAGAAACCTGTCTCAGAAACACACACAAACCAACAAACCAAAACCAAAACCAAAACCAAA
AAATAAATAAATAAATGAAGTGAAGAGCTTTGCTTATCAGAGATTTGTTGTGAGCATGTGAAGTGGACAGACTCCAGAAAGATGGGATTTTGA
CTCAGACACAAACAGCCCACTGACAAAGATGAATTTCCCAATTAGAGTATGTGGCATGAATAGGACATCAGAGAAATTTATATAGTGTCA
CAAAATATAAATAATTTTACGCCCACTTAGTAACTATGATAAATGCAAAATTAATTAATAGGATATATCTTACTATCTAGATTTAGTCAGACT
AGATGGCCATCTTATGTTTGGCAGCAGCATGATAGTAACGGAGTGTCCACATACAGAGATGAATGTTATGATCCCACTTATGACAGTGC
TAAATGTGTCTCAAAGTTAAACATACCCATGCTCTAGAGATGAGCATCTTAATGTCTTAGGACATGCGCAAGAAATGTCATGTGACCAAGACAG
ATACATGACAAATACATACAGAGCTCATCATGCAATAGCCAAAGTTAGAGACCAACAAATGATTACTGTGTATAGGAATGAATTTGTGATAGT
TGACAGGCTGATACACTTTTAACTTTGGGAGGAGCAGAGCAGGCACTCTGTGAGTCTGAGGCGCAGCTGGTCTACTTAGTGAATTTGGGAC
AGTCAAGGCTATGGAAGAGACTGTCTCAAAACCAAAATAAAACAAACCAACAACTAGTGGCAGTGTATGTTGGTGGTGTGTGATAGTGA
GTGGGCACTGTTGAAATAGAGTGTGTTTACTAGATATTTGACCCAGCTGCTCAATTTGCTCCAACTATGATGAACCCCAAACTCAAACTGGG
CACTGAGATTTAATTTCAAGGGTTGATGAACCACTGCGAGGCTGAGCACTTAGTTCAGTGTCTGGGACCACTGATGGTGAAGAAAGAACAA
CTGTTATAGTGTCTCTGATGCTCTGCTGTGATGATACACAAAGAGAGATTAATAATATAGTCTGTGAAGAAACAAATAAATCAATTAAGTGG
CCGCACTGCGCAGGCTTTAAACCCAGTACTGGGAGCAGAGCGAGTGAATTTTGTGAGTCTGGAGGCGAGCTGGCTCTCAGAGTGTGTCAGG
ACAGCGAGGCTACACAGAGAAACCTGCTTTGGAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAA
AGATAGTCTTCTAGTGAAGAACAGGAGTATGCTGCTTGTGAGCGAGAGGCTACAGTGGTGAAGTCTTATGGTTCTGAGCAACTGGTATTTGCC
TGGGTGACTAGTGTTTACAGGTATGTGCTTTTAAACAACTGTCAATGCTGCTGTGCAATAATACAGAAACCTGTGCTCATATGAAGAA
GTAAATATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG
ATACATACAGACGCTGAAATGCGACGACAAATCTGCTATGCAAGCAGATTTAAATAATTTAAATAAATAAAGAACCCAGAGATCTT
CTCATGCTGTAGAAACCACTCCAAGCTTAAGATCTGTCTAGGATGGAGCAAGTGAACAGTGAATTTAAATGTTGTAGTGTCTCATTTGTT
GGTGTGGACACTAAAGCCTTTTAGTGTAGTGAAGTGAATCATAGTACTGATATGAGTATGAAGAAGCAGGAGGACACTGCG
CTACCTGAAGTAAATATGCTGAGAACTTTGAATAAATCTGATCTGCTTGGCATGGGATACATGCTCTCTTCAAGTGTGAGGCTGATCTCT
CAGAGTCAAACTCTGAGCTAGTATGATGAGTCAAGTCAAGTCCCTAACAGAGGACAGGACTGAGCTCAGAGCCCTGAGCTGCTTTTT
CCCCGCTCTGAGCTTTGTTGATCCCTCCAGGCTTTACCCATGACGAAACCAAGGACCATCAGACCCCTAAGCATAGTGGCTTACAGACAGGAC
CAGCAAAACACTTGAAGCTTGAAGCTTTAACTTTAACTCTAGGAAGGAAATGTCTAGAGGAGGCTGTAGCAGGCTGTAGCAAGTCAACAGCA
GCAAGGTTCTCTGTGAGAGTCTGTGAGCTATCTGTGCCCAACACAGTATCTTTGTGCTATTTCTAGCTGGGGTGGGTTGAAATATGTAAT
GGGATTTAGTGAAGGTTTAACTATGAGCAATGGGAATAGGAACTGGAATCTGTATTTAGTCTTGGGAGCAGAGCTCTCTGAGCTGCTG
TCGCGTACAGGCTGCAATTCACACTTGGATTTTCTTCTCCATGTAACTCCCATCTCCATGCTGTGCTGCTCTCATCTCTCTCTGTT
TGTCTCTCTCCCTGAAGATCATCATATCTCAGCTGTCCAAAGTATCCCAATGGAATGCACTGAGCTAAGTCTCTGCTGCTCTCTGCTGCACT
GTTGATTTGTTAGTGTCTGACTTGTGATGGGCACTTCACTGACTTTCCACCTCCCCCTCAGTGGGAGCGGCTGTCTGTGTAGACTGTTGT

775

[illegible]

ATGTCCGTCGAGGCGCCCTCAGTCTCTCCACGACGAGCGCGTCATCAAAGCTGTCCCTTTCTCTCAACCCGACGGCTAACTTTGAAGGAAGTTTTTGA
GAATGGGAAACCTAAATGGATCTTTTGA AAAACCATTTAGTAAAGAAGGTGCGGTGGAAGAGGAGTGGCTTAAAGATCATCAATGATGGGGCTG
CCATCCTGAAGCAGGAGAAGACCATGATAGAGTGGAGGCTCGATCACAGTGTTGTGTGATGTTTCATGACAACTCTTTGACCTGATGAAGTTGTTT

[illegible]

HUMAN SEQUENCE - GENOMIC

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

126

[illegible]

128

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

GGGACACCCCTTAGCAGCGGTCGCGGTGCGTGCCGAAGCGGTGTTCCCGCGCTTAGCCGCTGCGCCTCCCAAGAGAGCGGCGGTTGGGCCCTCGTCTGTG
TCAGTGGCGTCGGAGGCCGCGCCTGCGGTGGCGCGCCCTTCTGGTGCTCGGACACCGCTGAGGAGCGGGGCCGGGCACGGCTGGCTGACGGCTCCGG

[illegible]

HUMAN SEQUENCE - CODING

[illegible]

BNSDOCID: <WO 03053224A2_1_>

138

139

140

141

142

143

144

146

[illegible]

150

151

152

153

154

CCACTATGGAAAAAGCTGGGAGATTCCCTTAAAAACTAAAGTAGATCTACCACTTTGATCCAGCAATCCCACTACTAGGTATTTACCCAAAGG
AAAGAGAGCTCATTTATACAAAAAAGAGTACTCTGCACAGCATTTTTATAGCAGCACAATTTGCCAATTTGCAATATATATGGAACCGAGCCCAATATGTC
CGATCAATCGAGTGGTAGGTAAGAAGAAATGGTATATATATATATACATACATATATATATATATATACACATATGGAATCTACTCT
TAGCTATGAAAAGGAATGAACAATCATAGTCTTTGCGACATCCTGGGATGGAATTAGAGACTATTTATCTAAGTGAGTAACTCAGGAATGTAAAAAC
CAACAATTTGATGTCCTCATCTAGTGGAGCATGAGCTATGAGGATGGAAAGGGCTAGAGAATGATCAACATGGAATCTTCAGGACATAGGGGAA
AGATGAGTGGGGGTGAGAGATGAAAAAGCTGCAATCTTTGGACAGCGCTGCTGCTGACCTGTTAATCCGACCACTTTGGGAGGCGCAGAGGTGG
CGAGATTCACAGAGCTCAGAGATCAGAGATCTGCTGACAGGTGGAACCGCTCTCTACTATTTAGGAAATTTAGCGAGGCGTGGTGG
CAGGCACTCGCTGCTGACCTACTCGGAGCGCTGAGGCGAGAGATGGAGGACCTTTGGAGGCGGATTTTGCAATGACGAGCTTGGGACAC
TGCACTCCAGGCTGGGCGACAGGCGAGACTCCCTCTCAAAAAAAGAAAAAAGAAAAATGGCTCTACCCCTAAGAGAGAGAAAGCGGCG
AGAGGAGAGAAACAATCAAGAGAGAACTAAGAGCAAGCAAGTGGAGCTTTAGCGCTTAGGCGCTTAGGCGCCAGAGGAAGCATTTGCTCAAAAG
AGCCAACTCTAGGATCTCTTTTTCCTTGGGGAACAGAGAGGGGTGAAGAAAATCGGAGTCAGAAAATTTGCTCCTGCTCCGCTCGAGGTAGC
GGGAGCTCCAGGTGATAATTAACAATCTTTAAAAAAGAAAGAAATTTATTTAAAAAAGAAAGAAAAATGCTGTAATTTCAATGGGCTGGG
GGGAGGGGAGGCTCTCTGGGATTAACAACCTGTAAACCTGCTAATCTCTTTTTCAGGCAATTTACACAGTGAATTTGAAGACAGTCTCAAGAGAT
TAAGTGAATTTACTAGAAATCTGATGATGATCTCTAAGACCAAGGCATGTAATTTTCCATATTTATTTACTGTCATTAGATCTATAAAAC
AGATACTACCATGTATACATTTTATCTTTTTCAGAAATTCAGAGATTTGCTTTAACTATCTGTCAGACAGCCAGAACAAACAGGAGAAAC
GAAACCAATTTAAACCTGTGGAATCTAAAGCTGAAACTAGTAATTTCTTTAATGAGGAGAGATGAGAGAGCAACAGAGACAGGGAAGCAAT
CAGAAATTTAGCAATPAGCAAGAGCTGCTCTCACTACCATAGGCTATGTTGAAAGGGGGAAGGAGGCTTAATGGAGCCAGGAGATTAAGTCTGAT
TGACAGACCTGGAGAACCAAGCTGGGCTGCTCAATTTGGTATCTGGAGCCGAGGAAGAGATACATCTTCTGTTGAGAGGCGCAAAAGATATAGAGC
ATAGGAGGAGAAATACCTTCCCTCTCCCTCCAGTCCACCCCTCTAGTGTTCCTGCTCTGCTTCTTATGGCTGTCACTGTTTCTAGGCTATACAG
CCAGAGAGCTGTGCTTCTGTCACACAGAGCAGAGAGAAAAAGTGGAGATGGATCTGAGAGATGGATCTGAGAGAGCAACAGAGACACAAAGCACCA
CACCTTTGTGATTTGGTGGTAAAGCTGGGATATAATACAGTTGACTAGATCTATAACAGACTGAAGAGCTAGTAGTGTGTGATGTTGATGTCAGGAGCA
ACGAGAGCTGCTGCTGGCTCCAGTCACTACAGCTGGAGACAACTCTCCAGGTGTGAAGGGTCTGTATTTAGTGTGAAGAGATAGCCAACTTAAC
TGAGCAGCAGATCATATTTAGCAAACTGATCTTATGACCTTAGGTTACATTAACAAACCTGGATATTTTGCTGCTCTGGATATATATCAGAA
TATAGATTTGTTGCTATCTATCTTTTAAAGACCTGGCAAACTTAACATATAACCAAGACTCATCTCTGGAAATGGGCTGAGAGATCTGATCT
TGTGAGATTAAGTAAAGAGCTGAGCTGCTTTTATGAGGCTGGAAGAGCTGAGAGGCTCTCCCACTATATATTTCAAAATTCAGAGAGGTT
ACTCGCAATATTTCTCAACATAATTTAGTGTGTTATCTACTCTTCAGACATCTACTAATTAATTAACAGTATAGAGATGATGATGATCT
TCCATTTTCTCAACAAAGCTGAGGTTTAAAGGCTGCAATPACCTAGCCCTTGGTCAAGGCTCTCAAACTCTGAGATCTCTCAACTCAT
GTGTTTAACTACGCTGTTACCAATATGTTGTGTTCTAATACTAGACAAATPCCAAATGGGTAGAGCTCAAGAGAGCCAAATTTGCTCTGATCAT
GAGAAATATCTTCCAAATGAGGTTAAATCTGCTATTGCTATTAATCATGSCAAAGACTGCTCTAGTGAGAAAGTAGGAGTTTCCATTTCCAGATG
TTTTCAGGCAAAAGTTTATTAGCAGTTTACTAATACTAGAGAAAAATATTACTAAAGATTAACAAAGGACCACTCTAATATTAAGAGGTGAG
TTTTAAAGGACCAATCAATCAAACTTAAATATGTTGCTCAATTAAGATATAGCTTCAAAATTAATTAACCAAACTTGGCAGAGCTTAGGAGAACCA
CTAAATCTCACAATTTACATTTAGATATTTCAACTACTCTCTATTACAAGTATAGAACAAACAAATTAGAAATATGAAACATCTGAAACCAACATCT
AGCCATCTTGAGCTAATAGACATATAGATATATAGAGACATACACCCGCAACTACAGAACACATGTTTCTTCAAGGCTATGTGGAACCAATCATCT
AAAGCAATTTACTAGAGGAATTTGCTCTAGAGAGGTGTTTTTCAATGAGCAATAAGTCAGATGATGAAATTTTTTGGATGATCTCCGCTCTTAAAT
TCTCAAAATTTACTCATCTTTCTGTAAGATATACCCCTCAGTCCAGCTTCCCACTGTGGAGAGGAGCTCTCTTTTCCACCACTTCAAAATTA
GAGAGAGGCTGTGGAGACATTTCTGTACTGACGAAAAATGTAGGATGTGGGCTCTGCTCCCTCCCAAACTCTGCTGATCTTTGTAGTGAAGAA
GCTCTCTCTGTCCTCATCTGCTCCCTCATGTATCCCAAGTCAAAAGCTCAGCCAACTCAGCCAACTCTGCGCATCTCTACTTCTGAAACCA
CTTCTCAACCGAGCAACCAAGATCACTTCTGCTGAGCAAGCACTGCAATAGTGGGCTGGACCTCAGCCACAGTACTCTGCTCTGTTCTGTT
TCTACTCTTCTTCCCAAAAGCTTGAGAGCTTACAGAAATAATCTACAGTATGATCATATAAAGATGATCAAAAGAGCAATGAAACAGGAA
AATAGATGTAGATATAATCAAAATGAGAGATAAAGTGAGTACTCGAAATACACACTCTGGGTACTGTTTGAATTAAGAGATAGGCCAAAG
TTTTGGCTCTCAGAGAGATTTGGTAGTCTCAAGTCCCAAGTCTCTCAGTAGGCTGCTCTCAGTGTCTGAATCTCTTACTCTACTCTCTGGG
TGCTCTCAAGCACTCAGAGCATTTCTCTCATCTGCTTTGTTGTCAGCATCTCTAGACATCTGCAATTTACACAGAGACATCACTCTTTATGTT
GATTCCAATTAAGATGCTATTACAAATGATATAATTAATTTTACCCCTCAGGCTCTGCAATCTTAAATGATTTTATATATATATATATG
TTTTATGTATTTTATATCAATTAGGGGGATGATCTCAGAGCTTTTTTGGACATACAGGAGCTGTGCTGACGAGATCTGAGAGCGA
GAACTCTCAAGGGGAGCCAACTGCCAGGAGTCTGTGGATGTTTCTATCTGCAAGTACCCCTGCTCTAGTAAACAGATCATGTTGGT
TGTGTTGCTATATAGTAATCAAAATTTCAACTTTAACAATTTGAGACCCACAAACAGAGGGGACAGGAGGCGTGAAGAGAGCCGAT
GAACATCTCTCTCTTCCATAGATGATGAAGGACGCTGGCAAAAGCTCAACACATAGAAAACAATTAAGTAATCATACAGTCAAGTGTGTT
CTACATGTTTGTATCAATTTTTAATGTGAAATTTTTTTCACAAAGTTTGTACTATTTATATCTCCACAAATTAGAGAAAGAAATTTGATAT
CACACCTCTGCTCTCTCTCAAGACTACACTTACTTCAAAATTTGTATTTGAAATTTAATTTCTTAAAGAAATATTTCTTCAAGAAAGTGAAGCA
GTGCTCATCCCTCTCTCTCTCACAAATATGACAGGAGACTTTCACACTGCTGTTGATCTGATATTTATGTTGTTTATTTTATG
CTATATCTAGACCACTGGAAGGTTGTTGCTGCTCTCTAGTCTCTTGAATATCATGCTCTATTTACTTATTTTAAAGTTTATTAAGG
TGTAAATGTTTATTTTAAATTTTAAATTTTATTTTAACTGCAATATGATATTTATGGGGTACAATTTGATTTTGTATGATGATAGAT
TATAGAAAGAGTCAATCAAGCGGCGCAGTGGCTCATGCTGTAATCCAGCACTTTGGGAGAGCAGGCGAGTGTGATCACTGAGGTCAAGGAT
GTTCAAAACCGAGCTGAGCAACATGATGTAAGACCCCATCTCTACGAAATAGAAAAAATAGCCAGGACATGTTGGCAGGCTGCAATCCGAC
CTCTCAGAGAGGTGAGGCGAGGAGATCTGTGAACCTGGGAGGTGGAGGTGGCAGTGGCGGAGATCAGCTGTGCTGCTCAGCTCAGCTGGTGA
CGAGCGAAACCTCTGCTCAAAAAGAAAAAAGAAATAAATCAAACTAGTTAACTACATCAAGCACTGATTTTGTGTTGTTGCTCT
TGTTTTTTCTCTTCTTATCTCTCAAAACAGAGGAATGTAAAAATTTCTGCTTTTAGCAAAATTTGAATACAGGTGATTTAAGTGTGCTGATCAC
ACATGGCAATGATCATAAAAAATTTCTCTCTCTCTGAGACTGAGCAACTCTTCCCATCCGAGCTCTGTAATACCCAGTTCTTCTACTCTGT
TCTTATGGAGTCTACTTTTTAGATTTTCAAAATTAATGAGATCATACATATCTGTCTTTTGTGCTGGCTTACTTAACTTAGCATATGACCT
CTTCACTTCCCTCAGTGTGTGCAAGATGACAGATTTTACTTCTTCTTGAAGGCTGTAGATTTTCACTGTGATATATACAGCAATTTCTTTTA
TCTTCTCTGTGTGATGACCAATTTAGTGTGCTTCCACTCTTGGACACCTGTTGAGGTCTGGAATCAATGATTTCAAAATTTGATGATGATGATG
ATAGCTCTATAGCTATGAAACAAACACTCBACTCTCTATCTTCAAGTAACCTGCACAGAAATTTGGAGATATCAGCAATGCACTCTTATGAG
ATTTGATGATTAATTTATTTATCTTCACTGCAATGAAGTAGCAAGCTCTATAGGAATCAGAAATTAAGGTCTTATTTCTTTCTCTCTATCT
GTCCAGCTCTCTCAGAAAAATTAACACTCTTAAAGTTTCAGAAACAGATGAGGTAGGCTCAGCTGCTGAGGCTCAGGTATCTCTCT
GGCAGAGATATCTCTGGCCAAAGCTCAGGTTCTCTGCGCAACACAGGACCTCAGGAGCCCTGAGGCGCTCTACAGGTCCAGAGTGTCTGCGC
CTAAGAGCCACCCTGCGCAAACTGCTGCTGTTCTGGAATCACTGAGTGGGAGCTCACTTTGGCCAGCATGACCTTCAAGATAGA

156

157

158

GAGATTCTTAATGGCTTAGGCCATTAGAATCTCTCTATACATTATCCAGAGAAATCTGGCACTCAACACCATTTAAATGTAACTACCACTG
 AGTCCAAATTTCAAGCAATCAACTAAGCAAAATAGGCCATTCAACAGCTGCTCTAGCTAATACATTTGGTTTCAGCTAGCTTTTGTATATACAA
 ACTTACCAAACTTAGTGAACTTAGTAAACCACTTTATGTTTTATGATTTTATAGGATAACAATTTGGGCTGGGCTCAGTGGAGGCAACTCAT
 TGTGCTCTACGTGGTGTGAGTGGATCATTCATGCACTTTGCAGTCAGTTGGCAGTTGGCTGGCAGATGGCTGGCTCAGATGGCTCATTCAT
 ATCTTTCTTATGTGTCTGGCTGTGGTGTGTAAGAAATACATGAAGGCGTGATACCACACTGTATCATCTACCAACATAGAGTCTTAATCCAAAGACAGCAA
 AAAAGAGAAACCAACCAATGTGACAGTGCTTTCAAAGTCTCTGTTTGGATACATTTGTGAATATGTTTCATATATCCAAAGCAAGACATATA
 CCACTGCCAAAATCAATGTGTGAAGAAATACATGAAGGCGTGATACCACACTGTATCATCTACCAACATAGAGTCTTAATCCAAAGACAGCAA
 TGTAGTATATCAGGACATCTTAATTTGGGCCAGTTCATAGTTATCTTCTGTGCTCTGGGCTTAATACATTTAGAAATCATCTGCTTCCCA
 CTGCTGGTCTTCTTAAATGTTTGAAGTAAACAAATCTTAGTATCTTTTGAAGTAAAGACTCTTCTGCTGAGCTAGATCTCTGCTTCTTCCCA
 CTCTGGTATATTCGGGAGACTGACTCTAAATAAGGCCATGAGAGTGTGGGAGCATCTGGGCTGAGAAAGATAGGAGTCTTCTCTGCAAGGCAA
 GTGCTCATATGTGGGCTTATACAGGCTTTCAAACAGGAAGAGGCTAATATCATTTAGACAGAAATGAGAGGCTCTTCCCACTATCAGGGAAC
 TCAAGGAACTCTGGGAGGTTTCAACATCTTCAGTTTCACTCCAAATCCACCAAAATGTCCTCCATCTCTAGCTCAGAGTCTCTCTCTTCAAAT
 ACTGAGAAATCTTCCACATGTGGAGCTTACACGCTTGTGAATGATGATTTGGTGTGAATTTGGTGTCTGTAATATGATTTGAGGTGAAGGCTTAACTTGTGA
 TGTATTTAGAGGATAGAGGCTTCTTCAGGCGAGTCTTGAAGAGGCTCTTCCCTCTTAATATGATTTCTTCACTATCAGGGAAC
 GTTCTCTCTCTAGACGCCCTGTGCTGATCAGAACGAGCCATCCCACTTACCTTCTTAATCCACCTAGCAGAACAGATGATTTTAACTACAGCTATGCTCAATGCTATGGA
 CAGCTACTTGGCTCCAGAAATCTGCTCTTCTGTTATTCACCTTAGCAGAACAGATGATTTTAACTACAGCTATGCTCAATGCTATGGA
 ATAGATTTACTAGTGGCAAGAGTTCAGCTCTCAATTCAAACCAATTCACCACTACACAAATCAATCAATCCCAAGTCTCAAGTCTTGTGA
 ATCTGTTCTCTATGACCACTGATGATACTCAAGTGTAGTATCAGCCAGAGTCCAGTCAATAAAATGAGTGTATTTAACTATAGGACTTGTGA
 TAAAGATATAAGAACTGAAAGCAGATAAAAGTGGATGCCAAATATAGAATACAGACAGCAGCTACTCTTCTACACTATTGTCTCAGGCG
 ATGACATCTCAGAGACTAGAAACAGGAAATAATTTCCCTCCACCTTCAGTCCCAAGGCTCTCCATCTTCTAGCTATTACAGAGAGCTA
 CTGTCGAGGGAGGCTGGCAATGAGTGTGCTCAAAATTCAGCCCTGGCAGCATCAGACGACGAGGAGGAGGATGAGCTATAGAGTGTAGAAC
 GAGGATACATACCAACTGACTCTCTCAATCAAGATCTATCTGGATGAAGTGGAGACATATCCATTAGAGATATAAAATACAGAGCTATATAAT
 TCCATTTAGATGTAGCACAAGAAATTTGCTTAACTCTGTGGCATCTGACCTGTAAACAGGCGCAATGATATTTTGTGATTTGTTTATG
 AGAAATAATAAATTTGGGCATCTGCTTTGACCATGAGCATAGGACAAATAAATTTATTTATTTATTTATTTATTTATTTATTTAT
 TATTTATTTTGTAGAGCGAGTCTCACTCTGCTGCGCCAGGCTGGAGTGCAGTGGCAAGATCTCGGCTCACTACTAGCTCGGCTCCGAGGTTCA
 TGCCATTTCCGCGCTCAGGACTCCAGTGTAGGACTCAGGTTGTGCGCGCCACTGCGCCGCTTAATTTTGTATTTATAGAGAGCGGG
 TTTCCCGGTGTAGCAGAGTGTCTCAATCTCTGACTCTGTGATCTGCCCACTCAGCTCCCAAGTGTCTGGGATACAGGCGTGGAGCCAC
 CGCAGCCGGGCTTTAAGTACTATTTGTAGAGTTGTATCAAAATTTCTGGCTGCAATGAGAGAAAGCTGTAGCAAACTGGCTTAATTTT
 TAAAGATATAAGTGTGGCTTATGTAACTGAGCAGCAGCAATGTGCACTGTGGTATAGGGGCTTAAACAAATGCTCTCAGGATCAAAATGTCTCT
 TATATTTCTTCACTCTGTTCTCTTTTGTGTATGTCNTTCTCAGACAGATATCTGCTCATTAATCAAGATGGATGGCAATGATTTATTAAC
 CTCTCAGTATCAAAATTTGTTTGTGTTATCTTTGCACTCTTATTTGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT
 TGTGCTGCCCTCTCCAGAACCACTCACTGAAGAAAGAGATATGTAGTATATTCATCTCTGGAGCTCCAGCCAGACCACTGAACTGAAAGT
 GGAAGTGAATCAAAAGTACATCTTTTGTGATAGTGTGCAATTTGCCCAATGCATCTCTGGAGCTCTCTACTAGAAGGATATAGCAGAAGTGT
 TAGGTGTCTATTATAGAAGAAATATCACTATCATATAACCCCTGGGGACATATTTCTAAATTT

HUMAN SEQUENCE - mRNA

GCAAAACCTTTAAGCTGAATGAACAACTTTCTCTCTTGAATATATCTTAAAGCAAAATTTGAGTGTCTTTTGTGATCCCATCTCATATGTCC
 CAGCTGGAAAGAAATCCCTGGGTTGAGCTACTGCTATGTGATGTTTGTGTTTTCTTTTGGCTGTCTCATTTTGGTGGCTACTATAGGAATCT
 AACCAAAACGCAAGCTGTTTTTGTGTTGTTATCTTTGCACTCTTACTTGTGGAGCTGTGGCAAGCTCATATCAAAATACAGAACTATGATCTCT
 CTCTCTGCTAATGTTGAGCTGGAAATGCACTTCACAGATAGCAGCTTTATTCACAGTGAACAGTCTCTAAGGAATCTGTATCAATATAGAGATG
 GCAGCAATGTGACCTTGAAGTCACTTTGACACTGGAAGTCACTGTGAACCTTGGAGCAATAACAGCAAGTTTGAAGAGTGGAAATGATATC
 ATGCCCAACACCGTGAAGAGCCACTTTCTGGAGGAGCAGCTGCCCTTAGGGAAGGCTCGTTCCACATACCTCAAGTCCAAGTGGAGGAGCAAA
 GAGACAGTACCAATGCATAATCATCTTGGGCTGCCCTGGAGCTACAAAGTACCTGACTCTGAAAGTCAAAGCTTCTAGAGAAATATAACACTC
 ACATCTTAAAGTTCCAGAAACAGATGAGGTAGAGCTCACTGCCAGGCTCAGGTTATCTCTGCGCAAGATATCTCTGGCAAGATCTCTGGCAAGCTCAGCGT
 TCTCTGCAACACGAGCCCTCCAGGACCTCTGAAGGCTCTACAGGTCACCAAGTGTCTGCGCTTAAAGCAACCCCTGGGAGAACTCTCAGC
 TGTGTGTTCTGGAATAGTCACTGTGAGGAACTTACTTTGGCCAGCATTTGACCTTCAAAGTCAGATGGAACCAAGCACTCTGCTCAAGTGTGCTGCT
 TTCAGTTCTTCTGCTCTCTGCTCATCTGCTTTTCAATTTTCATAGCCAGCATGATAGCCCTAAGAAACCACTCTGTCTCAAAGTCTTGTATTTCTC
 AAAGACACAAAGAAAGCTGTCCACCAACAAAGAGGAGAGTGAACAGTGCTATCTGAACCTGTGCTCTGGAGACAGGTTGACTGTATCTGTA
 TGACATCTAAGAGCTCTTGTGAATCTGGAACAAATTCGCTGTGCTGAGAGCTTGCATTTTCAATTTTCAATTTTGAATGAGCCAGC
 A

HUMAN SEQUENCE - CODING

ATGATCTCTCTCTGTAATGTTGAGCCTGGAATTCGAGCTTACCAGATAGCAGCTTTATTACAGATGACACTCCCTAAGAACTGTACATATA
 TAGAGCATGGCAGCAATGTGACCTCGGAATGCACTTTGACACTGGAGTCATGTGAACTCTGGGCAATACAGCCAGTTTGCAAAAGTGTGA
 AAGATGATACATCCCACACCGTGAAGAGGCCATTTGCTGGAGGAGCAGCTGCCCCTAGGAGAGGCTCTGTCACATCTCTCAATGCTCCAGTGT
 AGGAGCAAGAGCAGTACCAATGATATATCTATGCGGTGAGCTGGGATCAAGTACTGAGCTCTGAAGTCAAGCTCTGACAGGAAAT
 TAAACACTACATCTTCAAGGTTCCAGAAACAGATGAGGTAGAGCTCACCTGCCAGGCTACAGGTTTCTGCTGGCAGAGATATCTGCGCCAA
 CGTCAGCGTCTCTGCAACACAGCAGCACTACGAGCCCTGGAAGGCTCTACAGGTCAGAGTTGCTGGCTGAGCCACCCCTGGCAGGACAG
 AACTTCACTGTGTGTGTTCTGGAATCATCAAGTAGGGAATCTACTTTGGCCAGCATTTGACCTTCAAAGTCAGATGGAACCAAGCACTCTGCTCA
 CTGGCTGCTTCAATTTTCACTCCCTCTGCTCATCATGCTTCTTCAATTTATAGCCACAGTGAATGAGCTCTAGAAACCAAGCACTCTGCAAGGCT
 GTATTCTTCAAAGAGACACAAAAAGACTCTGCACCAACAAAGAGGAGGAGTGAACAGTGTCTATCTGA

MOUSE NOMENCLATURE	
ICSGNM	N/A
Celera	mCG2257
HUMAN NOMENCLATURE	
HGNC	PRDM11
Celera	hCG2538

[illegible]

162

163

164

165

166

167

168

170

470

AGGACAGAGAACATGAAGGATGCTGGGCCACACCAAGGCGGCTGATGGGGACATGGTGAAGTGAAGAACAAGAGTCTGCTCCACCATCTCC
TGCTGACGAGGATGGCGGACCGCTGCTTCGGAGATCTGGAGCATCTTCTCAGTAAGTGTGAACCCAGTAAGAATCGAAGGAGAGCGATCATCT
TTGTGACCAAAAGCTCTCCACAACTGGATCTTCTGGTCTTGGGCTCTGCGCAAGATGATCTCTGGTGAATGATGACCCCAAGACAGCTCCCCCGGTG
TTTGTGTCTGACAGCGACAGCTCTGTGGGACATCCAGATGGGCGAGCCCTACATCTTCCAGGCGATGGAGTGGTAAAGAGTGTCTGGCGGG
AGAGCGACGTGGCGGTGTATAAAGGAGTCACTCCCAAGGCGCATCTTTGGCCCTCTGAGGGCGAGATCTCTCCACCAAGACAGATCAAGCTCT
CTTCTCTTCACTGGGTGATTTGTGGACAGAACACCGCTCAAGTCTCAGATGGTGGTGAAGAGACCACTCAAGTAGAGTATCTGTGGTCTG
ATTCTCGGAGAGAGAGGAGGAGCAAGCACTCTGGGCTCTGCGACAGCGAGCGATCTTCTCTGGGACATCGCGAGACATCGCGCAGAGAGCG
GCTGAGAGAGAGAGGCTGAGACGATGACATGAGAGCGCTCAGTACGATCTCCAGGAAGACATCCAGCCAACTTAGCCGGGGAGAGAGAG
GCTCGAGAGAGAGAGGCTGAGACGATGAG
TACAGAGCGAGCTTTGATGGGCGAGACATCACTCTCAGGACAGAGAGAGAGAAATGATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
AGCTGTGGATCTGTGAAGCGCCGACAGCTGGCCCTGAGACCTCTCTGGTATCTCGGAAGATCCCAAGATCCAGGACAGATGATCTATGTAGAG
TGCCCTGTGACGAGAGCATCTGCGAGACCTCGGGGGAGAGGAGCTGGAAGTCTCTCAAAGAGCTGGCCAAAGAGCTGGGCGCATCGGAATGAA
GAGGAGAGAGCTACATATTAAGGCTGACAGCTCTCAGAGGCGCTCTTGCGATCTGACCCCATAGAGCTGCTCATACCTCTTTTGTCTCAT
ACTGATCTCTGCGTAAAGAAAGAGTGGCGGAATCTCAGGCGCGAGTACTAGCATGTCTTAAGTCTGAAAGAGCTGCGTAGGCCCTCCTTCTCGCCAC
CAGGTGTCTGAGCTCCAGAGTCTCAGTCTCTGCGAGTAAAGTCTTCCAGATGGAGTGTCTGTAAAGAGAGAGAGTGGATCTCAGCATGAA
TCTGATCTCAAAGCAGTGATGAGGCGCAACCGATGACCTCTCAGATGTCTCTTCTCGACAGAGCGCGCATCTCGGCGCTCTAAGCAGAGATGGC
TGAAAAGTCTTGTTGCTCTCAGCTTATCTCCGACAGCTGAATGAGATGTGGTGTCAAGTGTCTGCGCAGATCCAGGCTCAGTCTCAGCTCAT
CGCTTCTATCATCGGCTCGACAGTGTTAAGATCCACCATCAAGTCTGACAGCAGACAGCATCTGCAAGAAATCTGTGAGAGTCTGTACAA
CTGTGCGATCTCTCCCGGAGAGAGTGTGGAGTGTGCGGACATGTGCTGTCTTCAACACGCCCTTACCACTCGGCGCTGGGCGAGGAGGCG
ACCTTGAGCTTCTGGGCGCTCTCGAGAGCTCTGCGGAAGTGGAAGCTCAAGTGTGGAGCATGATTAAGTAGAGGCGCTGCGCATCTCTTAT
CATCAACATCTGGGCGGCTCAGAGAGAGACTCTGGGAGAGAGAGCTCCGTCAAGTCTCGCTCAGCATCTCTGGATGGAGTGAAGAGAG
ATTCTCAGAGAGAGAGTCTTCTCGGAGTATCTGAGTATACCGGATGATGAG
GAGAGAGAGAGAGTCTTCTCGGAGTATCTGAGTATACCGGATGATGAG
GGATGGGCGCATCACTCAGCGCAGCTTGGCGCGCAGCTCTTCTGACATCCGCAAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
CAGCGCGCTCACTGGAGATCTCTGATGATGACATCTGTGGAAAGAGCTGCCCTCTCTCAGAGAGCTGAGAGACAACTTGAAACAGCTCTCTAT
TACGCGCTTACAGCAGCTCTCATGTCTGTGATCTCTCAGCTCAGGCGCTCAGCTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
GAGGTGGATCATTTGGTGAAGAGATGTCTGAAGCGCCCTTATACAGGATCACTTGAGAGTGTGGTGCATCTGTAAGATGTTCAGCAGCAGACA
CAGGCTGGCTGAGCGCTCAGCATAGTCCCTGGCTCTCAGTCTCTCATGAGTACCAATCATGAGCTCATCTCTCTCTCAGTCTGAGTGTCT

176

177

178

BNSDOCID: <WO_____03053224A2_1_>

180

181

182

183

AOCCATGGTGTGCAGAAATATAGGCCAGACCCAGGGGAGGGGGACTGGAAGTCCCCAGGGGGTCTCCAAGGAGCCAGGCCAATTGGAGGATG
AAGAAGAGGAGCCTTCATCATTCAAGGCCGACAGTCCGCGAGGCCCTCCCTTGTCATCTGACCCCTCATGAACTTCCCACCACTCTTTTGGCC
TAACTGTATTGCGCTAAAGAAGAAGTTGGGAGCTCCAGGCAGAAATTAGACATGCTTAAGTCTGGGAAACTTCTGAGCCCCCGTATTGCCA
CCAAGGTACTGAGCTCCCAGGTTCTGGACCTGCAGGTAAGTTGGTTTGGATGAGATTATTGTGGAGGGCAGAGTACGCAGTGGGCTGT
GTGGAGGGTAG

187

188

189

190

191

192

193

194

195

196

197

198

[illegible]

202

204

BNSDOCID <WO_____03053224A2_1_>

206

207

208

209

210

211

212

213

[illegible][illegible]

GGCCCTCGTCCGCTCTGTGTCGAAGGACTCTCAAGAAGCGCTCTCCCGCAGAGCACTGTATCACGAAAATCAAGCGGAATTTACGACAG
CGTGGCTGCTCTCGGGCAACACGACCCTCAACGGGAGAGATGGGTTGGAGCAGAGCGCCATCAAGGTGTCTCTGAAGTGCCCATCACATTTC
CGCGCATCCAGTCCCTGCTCGAGGACACGATTGCAAGCATGTGCATGCTTTGATCTGGAGTCACTACTGCAGCTGAATTGCGAGAGAGGGA
CCTGGAGGTGCTCTGTGTGCAATAAAACCGCTCTGCTGGAGGCGCTGGAGTGGATCACTACATGTGGGAATCCTGAAATSCCATCCAAACATC
CGAGTTTGAAGAGGTACCATCGATCCACGTGCAGCTGGCGGCGCTGGCCCATCAAGTCCGACTTACACATCAAGAACGACCTGATGGCATC
CCCTCCAAGCGGTTCAAGACCATGAGTCCGAGCAGATGATCATGCCAATGTCTATGGAGATGATCGCAGCCCTGGGCCCCGGCCCTGCTCCCT
ATCCCTCCCGCTCCCCAGGGGGCACCACTCCAAAGACTACAGCAGCCAAAGCAACAATACTCAAGGCCATGGCACTTTGACTTCGCCCA
CGGGAACCTGGAGGGACATCCATGAATGACTTCATGCACGGGCCCCCAGCTCTCCCAACCCCCGACATGCCCAACAACATGGCCGCGCTC
GAGAAACCCCTCAGGCCACCCATGCAGGAACATATGCCACAGCTGGCAGCTCTGACCAGCCCCACCCCTCCATACAACAGAGTTTGCACGTAC
CACACCCAGCAGCCAGTCAGGGCCTCATTACATCACAGTGGGGCTCCTCCTCTCCTTCCAGCCTCCCCGGCAGCCGCAACAGGCGC
TCCAGCAGCCATCCACACAGCGACTGACCTTTAACCCCTCCTCAGCCTTAGAGGGTCAGGCGGAGCGCAGGGAGCGTCCGACATGCCGAG
CCTTCGCTGGATCTCCTCCGAACCTCACAAATCCTGACGAGCTCCTGTCTTATCTGGACCCCCCGACTGCGGAGCAATAGTACGATGACC
TCCTGTCTTATTGAGAACAACTGA

MOUSE NOMENCLATURE
ICSGNM Lfng
Celera mCG14497

HUMAN NOMENCLATURE
HGNC LFNG
Celera hCG18436

[illegible]

[illegible]

219

220

[illegible]

MOUSE SEQUENCE - mRNA

[illegible]

MOUSE SEQUENCE - CODING

ATGCT CCAAGCGGTCGGCGCGCGGCTCTGCTGGTGGCTGTGGGCGCTGTGGTGTCTCTCTGGTGTCAAGCGCGACCGCCGACCACTC
CGATCCGCGCTGAGCGCGACCGCGCGCTGCTGATGACGTCGCGGCGCTCTCGTGAGAGACTCGGCTCAGAGGTCCAGGCGCGTCTGGATCG
CGGAGTCTCCACCGCGAGGTGATCGATCTCTCGAGTACTCAGTCTACTCACTCGCGCGCGCAGAGAGCGGAGTACACCGCGCGGGTGGT
TCTCGCGAGGCGACGGCATCGGTCGCCCGCGGAAGTCTCATCTCTCGCGAGGCTTTCATCGGCTCAGAGCAACAGAAAGTTTCAAC
GCGCCGCGGCTCGATCTGTCTTGAGCAAGTGATCTCGCGCGCAAGGAAATGACGTTTCATCTCAGTATGGGAGAGCAAGGACTCTGGCCAA
CTCAGCAGGCAATGTGGTGTGACCAACTGCTCTCGCGCGACCGCGCAAGCTGTGCTGAGGATGAGGCTGTGGAGTATGACCAAGTTCATT
GAGTCTGGGAGAGAGTGTCTTGCCGACGCTGGATGATGACCAACTACGTCACACTCGCGGCGCTCTCGCGCTCTCGCGCAGTATCCCAACCG
AAGTGTGTGATCTCGCGAGGCGACCGCTGCGAGCGGCCCTCAGGCGCGCAAGAGTACGCGAGCAAGAATGAGACTCTCCATCTTGGTGT
CGATCCGCGTCCGCGAGTGCAGTGCACATGTTGCGGCTGCGGCTCAAGATGGCGCGCCCTGAGGCGATCTGGAGGACACTCTATGAGCAGCAGAG
TAGAGAACCTCCGCGAGTGCAGCAGTGCACATGTTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT
CAGGCGAGGACATCTCTGTGGAGCTGACCCATCAGGTTCCGCTCTGTCCATTGCCAACTGTAACCGAGACACACCTTGTGTCTCTGCTGAGT
ATCTCTATG

HUMAN SEQUENCE - GENOMIC

[illegible]

222

223

224

[illegible][illegible]

ACCTGGAGAACCTGCAGCAGGTGCCACCTCAGAGCTCCACGAGCAGGTGACGCTGAGCTACGGTATGTTTGA AAAACAAGCGGAACGCCCTCCA
CTGAAGGSGCCCTTCTCGGTGGAGGCCGACCCATCCAGGTTCCGCTCCATCCACTGCCACTGTACCGGACACACCTTGGTGTCCCCGCACT
GCCATCTTCTAG

HUMAN SEQUENCE - CODING

ATGCTCAAGCCTGCGCGCGCGCTGCTGCTGGCGTGGCGGGCGCCTGCTCGCCTGCTGCTGGTGTCAACGCCGAACGCCCGCCGCTC
CACTGCCCGCGAGCGCGCGCGCGCGCTGCGCGGCGGCCCGCGCGGGGCTGCCCGCGCGCGCGCTGGGGCGCGCGCGCGCGCGCG
GCTGGGGCGCTGCTCGCGACCTGCAAGTCTGTCCGAGTACTTCAGCCTGCTCACCCGCGCGCGCAGAGATGCGGCGCGCGCGCGCGCT
GCCCG
ACCGCGCGCGCGCTCGACTGCTGCTGGAGACCTGGATCTCG
CAGGCACACGGGCAACCTGGTTCATCACAACTGCTCG
ATCGAGTCCGGCAGGAAGTGGTTCTGCCACGTGGACGATGACAACTACGTCAACCTGCGGGCGCTGCTGCGCGCTGCTGGCGAGCTACCGCGCA
CGCGGACGTCTACGTCGGCAAGCCAGCCTGGACAGGCCCATCCAGGCCATGGAGCGGGTCAGCGAGAACAAAGGTGCGTCTGTCCACTTCTG
GTTTGCACGGGCGCGCTGCTGCTTCTGCATCAGCGCTGGCTGGCTCTGAAGATGAGCCGCTGGGCGAGCGGGGTCACTTCATGAAACGCGT
GAGCGATCCGGCTGCTGATGACTGCACCATCGGCTACATCGTGGAGGCCCTGCTGGGTGTGCCCTCATCCGAGCGGCGCTCTTCCACTCC
ACCTGGAGAACCTGCAGCAGGTGCCCACTCAGAGTCCAGAGCAGGTGACGCTGAGCTACGGTATGTTTGA AAAACAAGCGGAACGCCCTCCA
CGTGAAGGGCGCTTCTCGGTGGAGGCCGACCCATCCAGGTTCCGCTCCATCCACTGCCACTGTACCGGACACACCTTGGTGTCCCCGCACT
GCCATCTTCTAG

CLAIMS

We claim:

1. A recombinant nucleic acid comprising a nucleotide sequence selected from the group consisting of the sequences outlined in Tables 1-10.
2. A host cell comprising the recombinant nucleic acid of claim 1.
3. An expression vector comprising the recombinant nucleic acid according to claim 2.
4. A host cell comprising the expression vector of claim 3.
5. A recombinant protein comprising an amino acid sequence encoded by a nucleic acid sequence comprising a sequence selected from the group consisting of the sequences outlined in Tables 1-10.
6. A method of screening drug candidates comprising:
 - a) providing a cell that expresses a carcinoma associated (CA) gene comprising a nucleic acid sequence selected from the group consisting of the sequences outlined in Tables 1-10 or fragment thereof;
 - b) adding a drug candidate to said cell; and
 - c) determining the effect of said drug candidate on the expression of said CA gene.
7. A method according to claim 6 wherein said determining comprises comparing the level of expression in the absence of said drug candidate to the level of expression in the presence of said drug candidate.
8. A method of screening for a bioactive agent capable of binding to an CA protein (CAP), wherein said CAP is encoded by a nucleic acid comprising a nucleic acid sequence selected from the group consisting of the sequences outlined in Tables 1-10, said method comprising:
 - a) combining said CAP and a candidate bioactive agent; and
 - b) determining the binding of said candidate agent to said CAP.
9. A method for screening for a bioactive agent capable of modulating the activity of an CA protein (CAP), wherein said CAP is encoded by a nucleic acid comprising a nucleic acid sequence selected from the group consisting of the sequences outlined in Tables 1-10, said method comprising:
 - a) combining said CAP and a candidate bioactive agent; and
 - b) determining the effect of said candidate agent on the bioactivity of said CAP.
10. A method of evaluating the effect of a candidate carcinoma drug comprising:
 - a) administering said drug to a patient;
 - b) removing a cell sample from said patient; and
 - c) determining alterations in the expression or activation of a gene comprising a nucleic acid sequence selected from the group consisting of the sequences outlined in Tables 1-10.

11. A method of diagnosing carcinoma comprising:
 - a) determining the expression of one or more genes comprising a nucleic acid sequence selected from the group consisting of the sequences outlined in Tables 1-10, in a first tissue type of a first individual; and
 - b) comparing said expression of said gene(s) from a second normal tissue type from said first individual or a second unaffected individual;wherein a difference in said expression indicates that the first individual has carcinoma.
12. A method for inhibiting the activity of a CA protein (CAP), wherein said CAP is encoded by a nucleic acid comprising a nucleic acid sequence selected from the group consisting of the sequences outlined in Tables 1-10, said method comprising binding an inhibitor to said CAP.
13. A method of treating carcinomas comprising administering to a patient an inhibitor of an CA protein (CAP), wherein said CAP is encoded by a nucleic acid comprising a nucleic acid sequence selected from the group consisting of the sequences outlined in Tables 1-10.
14. A method of neutralizing the effect of an CA protein (CAP), wherein said CAP is encoded by a nucleic acid comprising a nucleic acid sequence selected from the group consisting of the sequences outlined in Tables 1-10, comprising contacting an agent specific for said CAP protein with said CAP protein in an amount sufficient to effect neutralization.
15. A polypeptide which specifically binds to a protein encoded by a nucleic acid comprising a nucleic acid selected from the group consisting of the sequences outlined in Tables 1-10.
16. A polypeptide according to claim 15 comprising an antibody which specifically binds to a protein encoded by a nucleic acid comprising a nucleic acid sequence selected from the group consisting of the sequences outlined in Tables 1-10.
17. A biochip comprising one or more nucleic acid segments selected from the group consisting of a nucleic acid of the sequences outlined in Tables 1-10 or fragments thereof.
18. A method of diagnosing carcinoma or a propensity to carcinoma by sequencing at least one CA gene of an individual.
19. A method of determining CA gene copy number comprising adding an CA gene probe to a sample of genomic DNA from an individual under conditions suitable for hybridization.

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 July 2003 (03.07.2003)

PCT

(10) International Publication Number
WO 03/053224 A3

(15) International Patent Classification⁷: **C07H 21/02**,
21/04, C12Q 1/00, 1/68, G01N 33/48, 31/53, 31/567,
31/574, C12P 21/06, C12N 15/00, 15/09, 15/63

(21) International Application Number: PCT/US02/41776

(22) International Filing Date:
20 December 2002 (20.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/034,650 20 December 2001 (20.12.2001) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PI, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(71) Applicant (*for all designated States except US*): **SAGRES DISCOVERY** [US/US]; 2795 Second Street, Suite 400, Davis, CA 95616 (US).

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): **MORRIS, David, W.** [US/US]; 2481 Emerald Bay Drive, Davis, CA 95616 (US).

(88) Date of publication of the international search report:
4 September 2003

(74) Agents: **BASU, Shantanu et al.**; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304-1018 (US).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/053224 A3

(54) Title: NOVEL COMPOSITIONS AND METHODS FOR CANCER

(57) Abstract: The present invention relates to novel sequences for use in diagnosis and treatment of carcinomas, especially lymphoma carcinomas. In addition, the present invention describes the use of novel compositions for use in screening methods.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/41776

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/02, 21/04; C12Q 1/00, 1/68; G01N 33/48, 31/53, 31/567, 31/574; C12P 21/06; C12N 15/00, 15/09, 15/63
 US CL : 536/23.1; 530/300, 350, 435/69.1, 320.1, 325, 4, 6, 7.1, 7.21, 7.23; 436/63, 64

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U. S. : 536/23.1; 530/300, 350, 435/69.1, 320.1, 325, 4, 6, 7.1, 7.21, 7.23; 436/63, 64

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
 GenCore databases, WEST, Medline

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Database GenCore on STN, Accession number AF275818, YANG et al. 23 July 2000 (23.07.2000), 'A family of novel PR-domain (PRDM) genes as candidate tumor suppressors'. Direct Submission.	1
A	JIANG, G.-L et al. The yin-yang of PR-domain family genes in tumorigenesis. Histol. Histopathol. January 2000, Vol. 15, No. 1, pages 109-117.	1-19



Further documents are listed in the continuation of Box C.



See patent family annex.

Special categories of cited documents:	
* "A" document defining the general state of the art which is not considered to be of particular relevance	* "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
* "E" earlier application or patent published on or after the international filing date	* "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
* "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	* "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
* "O" document referring to an oral disclosure, use, exhibition or other means	* "&" document member of the same patent family
* "P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

08 May 2003 (08.05.2003)

Date of mailing of the international search report

21 JUL 2003

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US
 Commissioner for Patents
 P.O. Box 1450
 Alexandria, Virginia 22313-1450
 Facsimile No. (703)305-3230

Authorized officer

Alana M. Harris, Ph.D.

Telephone No. (703)308-0196

Form PCT/ISA/210 (second sheet) (July 1998)